Unix Practice

Joann Mudge

February 18, 2021

- Log in to logrus.
  - ssh -p 44111 logrus.training.ncgr.org
- Create a screen.
- You should be in your home directory.
- Make a directory called "unix_practice" and go into it.

```
mkdir unix_practice
cd unix_practice
# Hint: don't forget that you can tab out "unix_practice" once you have made it.
```

- Now make the following directories:
  - assemblies
  - expression
  - foo
- List the directories (hint: use the -l option in order to see the path to the linked file).
- Remove the foo directory (for directories you need to use the rmdir command rather than rm).
- Go into the assemblies directory.

```
## total 12
## drwxrwxr-x 2 jm jm 4096 Feb 18 09:06 assemblies
## drwxrwxr-x 2 jm jm 4096 Feb 18 09:06 expression
## drwxrwxr-x 2 jm jm 4096 Feb 18 09:06 foo
```

Go into the assemblies directory and make a soft link to the data. The data are big so we will link to it, which just points to the data rather than copying it.

```
cd assemblies
ln -s /home/jm/unixdata/assemblies/* .
```

Check the directory and path that you are in using the pwd command (print working directory).

```
pwd


## /home/jm/unix_practice/assemblies
```

Now go into the expression directory and check the path and directory.

```
cd ../expression
pwd



## /home/jm/unix_practice/expression
```

- Make a link to the data (/home/jm/unixdata/expression/*).
- List the files again.
- Go back into the assemblies directory.
- What size are the files?
- How many assembly files are there? They are a subset of yeast assemblies from this paper: https://www.nature.com/articles/s41586-018-0030-5.pdf
  - Hint: a pipe symbol (|) tells unix to send the results of one command directly into another command.

```
## total 0
## lrwxrwxrwx 1 jm jm 47 Feb 18 09:06 3.boxplots.txt.bak -> /home/jm/unixdata/expression/3.b
## lrwxrwxrwx 1 jm jm 53 Feb 18 09:06 boxplots.10highestFC.txt -> /home/jm/unixdata/expressi
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 expression.txt -> /home/jm/unixdata/expression/express
## lrwxrwxrwx 1 jm jm 37 Feb 18 09:06 expr.txt -> /home/jm/unixdata/expression/expr.txt
## lrwxrwxrwx 1 jm jm 39 Feb 18 09:06 rowsToCols -> /home/jm/unixdata/expression/rowsToCols
## lrwxrwxrwx 1 jm jm 40 Feb 18 09:06 run-fc.bash -> /home/jm/unixdata/expression/run-fc.bas
## total 0
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAA_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAA_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAB_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAB_6.r
```

```
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAC_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAC_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAD_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAD_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAE_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAE_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAG_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAG_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAH_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAH_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAI_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAI_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAK_6.re.fa.gz -> /home/jm/unixdata/assemblies/AAK_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAL_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAL_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAM_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAM_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAN_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAN_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAP_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAP_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAQ_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAQ_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAR_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAR_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAS_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAS_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAT_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAT_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AAV_3.re.fa.gz -> /home/jm/unixdata/assemblies/AAV_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABA_3.re.fa.gz -> /home/jm/unixdata/assemblies/ABA_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABB_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABB_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABC_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABC_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABD_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABD_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABE_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABE_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABF_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABF_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABG_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABG_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABH_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABH_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABI_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABI_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABK_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABK_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABL_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABL_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABM_5.re.fa.gz -> /home/jm/unixdata/assemblies/ABM_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABP_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABP_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABQ_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABQ_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABR_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABR_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABS_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABS_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABT_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABT_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ABV_6.re.fa.gz -> /home/jm/unixdata/assemblies/ABV_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACA_6.re.fa.gz -> /home/jm/unixdata/assemblies/ACA_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACB_6.re.fa.gz -> /home/jm/unixdata/assemblies/ACB_6.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACC_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACC_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACD_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACD_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACF_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACF_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACG_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACG_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACH_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACH_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACI_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACI_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACK_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACK_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACL_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACL_7.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACM_7.re.fa.gz -> /home/jm/unixdata/assemblies/ACM_7.r
```

```
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACN_8.re.fa.gz -> /home/jm/unixdata/assemblies/ACN_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACP_8.re.fa.gz -> /home/jm/unixdata/assemblies/ACP_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACQ_8.re.fa.gz -> /home/jm/unixdata/assemblies/ACQ_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACR_2.re.fa.gz -> /home/jm/unixdata/assemblies/ACR_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACS_2.re.fa.gz -> /home/jm/unixdata/assemblies/ACS_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACT_2.re.fa.gz -> /home/jm/unixdata/assemblies/ACT_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ACV_2.re.fa.gz -> /home/jm/unixdata/assemblies/ACV_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADA_2.re.fa.gz -> /home/jm/unixdata/assemblies/ADA_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADB_2.re.fa.gz -> /home/jm/unixdata/assemblies/ADB_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADC_2.re.fa.gz -> /home/jm/unixdata/assemblies/ADC_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADD_2.re.fa.gz -> /home/jm/unixdata/assemblies/ADD_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADE_2.re.fa.gz -> /home/jm/unixdata/assemblies/ADE_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADF_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADF_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADG_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADG_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADH_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADH_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADI_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADI_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADK_8.re.fa.gz -> /home/jm/unixdata/assemblies/ADK_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADL_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADL_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADM_2.re.fa.gz -> /home/jm/unixdata/assemblies/ADM_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADN_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADN_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADP_3.re.fa.gz -> /home/jm/unixdata/assemblies/ADP_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADQ_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADQ_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADR_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADR_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADS_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADS_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADT_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADT_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 ADV_4.re.fa.gz -> /home/jm/unixdata/assemblies/ADV_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEA_8.re.fa.gz -> /home/jm/unixdata/assemblies/AEA_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEB_8.re.fa.gz -> /home/jm/unixdata/assemblies/AEB_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEC_8.re.fa.gz -> /home/jm/unixdata/assemblies/AEC_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEE_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEE_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEF_8.re.fa.gz -> /home/jm/unixdata/assemblies/AEF_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEG_8.re.fa.gz -> /home/jm/unixdata/assemblies/AEG_8.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEH_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEH_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEI_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEI_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEK_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEK_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEL_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEL_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEM_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEM_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEN_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEN_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEP_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEP_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEQ_3.re.fa.gz -> /home/jm/unixdata/assemblies/AEQ_3.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AER_4.re.fa.gz -> /home/jm/unixdata/assemblies/AER_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AES_4.re.fa.gz -> /home/jm/unixdata/assemblies/AES_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AET_5.re.fa.gz -> /home/jm/unixdata/assemblies/AET_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AEV_5.re.fa.gz -> /home/jm/unixdata/assemblies/AEV_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFA_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFA_4.r
```

```
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFB_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFB_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFC_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFC_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFD_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFD_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFE_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFE_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFF_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFF_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFG_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFG_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFH_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFH_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFI_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFI_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFK_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFK_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFL_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFL_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFM_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFM_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFN_4.re.fa.gz -> /home/jm/unixdata/assemblies/AFN_4.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFP_1.re.fa.gz -> /home/jm/unixdata/assemblies/AFP_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFQ_1.re.fa.gz -> /home/jm/unixdata/assemblies/AFQ_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFR_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFR_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFS_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFS_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFT_5.re.fa.gz -> /home/jm/unixdata/assemblies/AFT_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AFV_1.re.fa.gz -> /home/jm/unixdata/assemblies/AFV_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGA_5.re.fa.gz -> /home/jm/unixdata/assemblies/AGA_5.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGB_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGB_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGC_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGC_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGE_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGE_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGF_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGF_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGG_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGG_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGH_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGH_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGI_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGI_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGK_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGK_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGL_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGL_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGM_1.re.fa.gz -> /home/jm/unixdata/assemblies/AGM_1.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGN_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGN_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGP_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGP_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGR_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGR_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGS_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGS_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGT_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGT_2.r
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 AGV_2.re.fa.gz -> /home/jm/unixdata/assemblies/AGV_2.r
## lrwxrwxrwx 1 jm jm 44 Feb 18 09:06 random_file.txt -> /home/jm/unixdata/assemblies/random
## 127
```

Take a look at the way the sequences are named. Then count the number of sequences in the file AAA_6.re.fa.gz. We will use zgrep so we can work directly on the zipped files.

```
zgrep ">" AAA_6.re.fa.gz | head
zgrep -c ">" AAA_6.re.fa.gz
```

```
## >AAA_6-0
## >AAA_6-4
## >AAA_6-7
## >AAA_6-8
## >AAA_6-9
## >AAA_6-10
## >AAA_6-13
## >AAA_6-16
## >AAA_6-17
## >AAA_6-20
## 4893
```

- Now count the number of sequences for the file AAB_6.re.fa.gz file.

```
## 5814
```

Find the total sequence length of the assembly. We need to generate a command that:

- removes the header lines so we don't count them
  - Using the -v paramter will tell grep to remove the lines that it finds.
- counts the length of each remaining line
  - We'll use a built in variable in awk to find the length of a field (we have only one field).
- adds them together
  - Again, we'll use awk.

We'll string all these commands together using pipes.

```
zgrep -v ">" AAA_6.re.fa.gz | awk '{sum+=length($1)} END {print sum}'
```

```
## 12091446
```

You can also get an idea of an assembly size by looking at the file size. It will be slightly off because it counts header lines and white space. But because there are generally far fewer characters in the header and in white space, it usually isn't too far off. Take a look at the file size using "ls -l" and compare it to the assembly length you just calculated.

Now take it up a notch. Count the number of sequences in each file.

```
zgrep -c ">" *.fa.gz


## AAA_6.re.fa.gz:4893
## AAB_6.re.fa.gz:5814
## AAC_6.re.fa.gz:4908
## AAD_6.re.fa.gz:2093
## AAE_6.re.fa.gz:2613
## AAG_6.re.fa.gz:2693
## AAH_6.re.fa.gz:1923
## AAI_6.re.fa.gz:3819
## AAK_6.re.fa.gz:1695
## AAL_3.re.fa.gz:5821
## AAM_3.re.fa.gz:2969
## AAN_3.re.fa.gz:7184
## AAP_3.re.fa.gz:6002
## AAQ_3.re.fa.gz:2782
## AAR_3.re.fa.gz:3060
## AAS_3.re.fa.gz:2389
## AAT_3.re.fa.gz:2449
## AAV_3.re.fa.gz:4577
## ABA_3.re.fa.gz:2557
## ABB_5.re.fa.gz:1682
## ABC_5.re.fa.gz:1723
## ABD_5.re.fa.gz:2401
## ABE_5.re.fa.gz:4239
## ABF_6.re.fa.gz:2363
## ABG_5.re.fa.gz:2145
## ABH_5.re.fa.gz:1866
## ABI_5.re.fa.gz:1924
## ABK_5.re.fa.gz:2156
## ABL_5.re.fa.gz:1509
## ABM_5.re.fa.gz:1483
## ABP_6.re.fa.gz:1994
## ABQ_6.re.fa.gz:1173
## ABR_6.re.fa.gz:1445
## ABS_6.re.fa.gz:3755
## ABT_6.re.fa.gz:1704
## ABV_6.re.fa.gz:20018
## ACA_6.re.fa.gz:4722
## ACB_6.re.fa.gz:2810
## ACC_7.re.fa.gz:2068
## ACD_7.re.fa.gz:1805
## ACF_7.re.fa.gz:1651
## ACG_7.re.fa.gz:4749
```

```
## ACH_7.re.fa.gz:7145
## ACI_7.re.fa.gz:2479
## ACK_7.re.fa.gz:2299
## ACL_7.re.fa.gz:2796
## ACM_7.re.fa.gz:3140
## ACN_8.re.fa.gz:3598
## ACP_8.re.fa.gz:2135
## ACQ_8.re.fa.gz:1281
## ACR_2.re.fa.gz:1662
## ACS_2.re.fa.gz:1459
## ACT_2.re.fa.gz:1399
## ACV_2.re.fa.gz:1343
## ADA_2.re.fa.gz:2436
## ADB_2.re.fa.gz:3076
## ADC_2.re.fa.gz:1384
## ADD_2.re.fa.gz:2254
## ADE_2.re.fa.gz:1654
## ADF_4.re.fa.gz:3299
## ADG_4.re.fa.gz:2790
## ADH_4.re.fa.gz:2919
## ADI_4.re.fa.gz:6983
## ADK_8.re.fa.gz:6118
## ADL_4.re.fa.gz:8801
## ADM_2.re.fa.gz:5934
## ADN_4.re.fa.gz:3703
## ADP_3.re.fa.gz:4169
## ADQ_4.re.fa.gz:3068
## ADR_4.re.fa.gz:3110
## ADS_4.re.fa.gz:3302
## ADT_4.re.fa.gz:2997
## ADV_4.re.fa.gz:2581
## AEA_8.re.fa.gz:2609
## AEB_8.re.fa.gz:10458
## AEC_8.re.fa.gz:13829
## AEE_3.re.fa.gz:9217
## AEF_8.re.fa.gz:1566
## AEG_8.re.fa.gz:2176
## AEH_3.re.fa.gz:3140
## AEI_3.re.fa.gz:5152
## AEK_3.re.fa.gz:2563
## AEL_3.re.fa.gz:7695
## AEM_3.re.fa.gz:5463
## AEN_3.re.fa.gz:6070
## AEP_3.re.fa.gz:5997
## AEQ_3.re.fa.gz:5626
```

```
## AER_4.re.fa.gz:5393
## AES_4.re.fa.gz:6438
## AET_5.re.fa.gz:3478
## AEV_5.re.fa.gz:6667
## AFA_4.re.fa.gz:3382
## AFB_4.re.fa.gz:3273
## AFC_4.re.fa.gz:4638
## AFD_4.re.fa.gz:3185
## AFE_4.re.fa.gz:5347
## AFF_4.re.fa.gz:3248
## AFG_4.re.fa.gz:3679
## AFH_4.re.fa.gz:5465
## AFI_5.re.fa.gz:2683
## AFK_5.re.fa.gz:3263
## AFL_5.re.fa.gz:2393
## AFM_5.re.fa.gz:6504
## AFN_4.re.fa.gz:2165
## AFP_1.re.fa.gz:2803
## AFQ_1.re.fa.gz:2056
## AFR_5.re.fa.gz:3359
## AFS_5.re.fa.gz:5267
## AFT_5.re.fa.gz:2574
## AFV_1.re.fa.gz:32895
## AGA_5.re.fa.gz:3024
## AGB_1.re.fa.gz:8475
## AGC_1.re.fa.gz:17216
## AGE_1.re.fa.gz:8269
## AGF_1.re.fa.gz:14304
## AGG_1.re.fa.gz:4793
## AGH_1.re.fa.gz:4547
## AGI_1.re.fa.gz:2613
## AGK_1.re.fa.gz:9979
## AGL_2.re.fa.gz:4846
## AGM_1.re.fa.gz:25588
## AGN_2.re.fa.gz:3062
## AGP_2.re.fa.gz:3469
## AGR_2.re.fa.gz:2150
## AGS_2.re.fa.gz:2260
## AGT_2.re.fa.gz:5158
## AGV_2.re.fa.gz:2887
```

It isn't as simple if we want to find the total length for all the assemblies. We'll need to use a for loop.

```
for i in *.fa.gz; do
printf "$i\t"
zgrep -v ">" $i | awk '{sum+=length($1)} END {print sum}'
done


## AAA_6.re.fa.gz 12091446
## AAB_6.re.fa.gz 12362047
## AAC_6.re.fa.gz 14102010
## AAD_6.re.fa.gz 11847030
## AAE_6.re.fa.gz 11941110
## AAG_6.re.fa.gz 12099456
## AAH_6.re.fa.gz 11898446
## AAI_6.re.fa.gz 12122078
## AAK_6.re.fa.gz 11719514
## AAL_3.re.fa.gz 14684741
## AAM_3.re.fa.gz 11971708
## AAN_3.re.fa.gz 14924236
## AAP_3.re.fa.gz 14998012
## AAQ_3.re.fa.gz 12029663
## AAR_3.re.fa.gz 12097491
## AAS_3.re.fa.gz 11955446
## AAT_3.re.fa.gz 12113089
## AAV_3.re.fa.gz 12184813
## ABA_3.re.fa.gz 12016484
## ABB_5.re.fa.gz 11831103
## ABC_5.re.fa.gz 11928376
## ABD_5.re.fa.gz 12039888
## ABE_5.re.fa.gz 12024093
## ABF_6.re.fa.gz 12045231
## ABG_5.re.fa.gz 11996940
## ABH_5.re.fa.gz 11978627
## ABI_5.re.fa.gz 12031849
## ABK_5.re.fa.gz 12022762
## ABL_5.re.fa.gz 11790281
## ABM_5.re.fa.gz 11780930
## ABP_6.re.fa.gz 11898606
## ABQ_6.re.fa.gz 11678037
## ABR_6.re.fa.gz 11789540
## ABS_6.re.fa.gz 11992627
## ABT_6.re.fa.gz 11879070
## ABV_6.re.fa.gz 16643749
## ACA_6.re.fa.gz 12182221
## ACB_6.re.fa.gz 11997694
## ACC_7.re.fa.gz 12026039
```

```
## ACD_7.re.fa.gz 12005874
## ACF_7.re.fa.gz 11790849
## ACG_7.re.fa.gz 12164522
## ACH_7.re.fa.gz 12389975
## ACI_7.re.fa.gz 12028650
## ACK_7.re.fa.gz 11911521
## ACL_7.re.fa.gz 12113657
## ACM_7.re.fa.gz 11971216
## ACN_8.re.fa.gz 12082117
## ACP_8.re.fa.gz 11864651
## ACQ_8.re.fa.gz 11926869
## ACR_2.re.fa.gz 11838854
## ACS_2.re.fa.gz 11904669
## ACT_2.re.fa.gz 11853592
## ACV_2.re.fa.gz 11853059
## ADA_2.re.fa.gz 11987860
## ADB_2.re.fa.gz 11967676
## ADC_2.re.fa.gz 11857431
## ADD_2.re.fa.gz 11873309
## ADE_2.re.fa.gz 11888642
## ADF_4.re.fa.gz 12073737
## ADG_4.re.fa.gz 12106138
## ADH_4.re.fa.gz 11979761
## ADI_4.re.fa.gz 12314491
## ADK_8.re.fa.gz 14336710
## ADL_4.re.fa.gz 15279666
## ADM_2.re.fa.gz 15279511
## ADN_4.re.fa.gz 12041283
## ADP_3.re.fa.gz 12271217
## ADQ_4.re.fa.gz 11938966
## ADR_4.re.fa.gz 11962626
## ADS_4.re.fa.gz 11970148
## ADT_4.re.fa.gz 11992679
## ADV_4.re.fa.gz 11945701
## AEA_8.re.fa.gz 12049442
## AEB_8.re.fa.gz 14127449
## AEC_8.re.fa.gz 15138865
## AEE_3.re.fa.gz 14289720
## AEF_8.re.fa.gz 11800789
## AEG_8.re.fa.gz 12048110
## AEH_3.re.fa.gz 12024296
## AEI_3.re.fa.gz 12268917
## AEK_3.re.fa.gz 11737259
## AEL_3.re.fa.gz 15291229
## AEM_3.re.fa.gz 12420411
```

```
## AEN_3.re.fa.gz 13422513
## AEP_3.re.fa.gz 12247265
## AEQ_3.re.fa.gz 12369422
## AER_4.re.fa.gz 12467493
## AES_4.re.fa.gz 12239172
## AET_5.re.fa.gz 11915610
## AEV_5.re.fa.gz 14757889
## AFA_4.re.fa.gz 12039160
## AFB_4.re.fa.gz 12035738
## AFC_4.re.fa.gz 14581796
## AFD_4.re.fa.gz 12077170
## AFE_4.re.fa.gz 13961355
## AFF_4.re.fa.gz 11936874
## AFG_4.re.fa.gz 12048776
## AFH_4.re.fa.gz 12130594
## AFI_5.re.fa.gz 11916104
## AFK_5.re.fa.gz 12028980
## AFL_5.re.fa.gz 11848436
## AFM_5.re.fa.gz 12245353
## AFN_4.re.fa.gz 11872363
## AFP_1.re.fa.gz 11960451
## AFQ_1.re.fa.gz 11972211
## AFR_5.re.fa.gz 12139182
## AFS_5.re.fa.gz 12802936
## AFT_5.re.fa.gz 11930374
## AFV_1.re.fa.gz 14634588
## AGA_5.re.fa.gz 11919684
## AGB_1.re.fa.gz 13125030
## AGC_1.re.fa.gz 13783091
## AGE_1.re.fa.gz 15029773
## AGF_1.re.fa.gz 12932748
## AGG_1.re.fa.gz 12401958
## AGH_1.re.fa.gz 12092775
## AGI_1.re.fa.gz 11978972
## AGK_1.re.fa.gz 12572138
## AGL_2.re.fa.gz 12043831
## AGM_1.re.fa.gz 14351270
## AGN_2.re.fa.gz 11964427
## AGP_2.re.fa.gz 12356070
## AGR_2.re.fa.gz 12023385
## AGS_2.re.fa.gz 11926963
## AGT_2.re.fa.gz 12354194
## AGV_2.re.fa.gz 11969461
```

Let's practice using screen.

- Create a new window in the screen that you are already in (ctrl-a + c)
  - You can have up to 10 windows
- Navigate between windows
  - Cycle from window to window using ctrl-a + space
  - Switch back and forth between two windows using ctrl-a + a
- Scroll up inside a screen
  - Change to scroll mode using ctrl-a + esc, then use the up arrow (or the down arrow if you scroll too far)
  - Get out of scroll mod using esc
- Exit the screen
- Re-enter the screen and go to the window you just made

- Go to the expression directory
- Look at the filenames

```
cd ../expression
ls -l
```

```
## total 0
## lrwxrwxrwx 1 jm jm 47 Feb 18 09:06 3.boxplots.txt.bak -> /home/jm/unixdata/expression/3.b
## lrwxrwxrwx 1 jm jm 53 Feb 18 09:06 boxplots.10highestFC.txt -> /home/jm/unixdata/expressi
## lrwxrwxrwx 1 jm jm 43 Feb 18 09:06 expression.txt -> /home/jm/unixdata/expression/express
## lrwxrwxrwx 1 jm jm 37 Feb 18 09:06 expr.txt -> /home/jm/unixdata/expression/expr.txt
## lrwxrwxrwx 1 jm jm 39 Feb 18 09:06 rowsToCols -> /home/jm/unixdata/expression/rowsToCols
## lrwxrwxrwx 1 jm jm 40 Feb 18 09:06 run-fc.bash -> /home/jm/unixdata/expression/run-fc.bas
```

Look at the first 25 lines of the file.

```
head -25 expression.txt
```

```
## Row Gene Function padj Sample1_Rep1 Sample2_Rep1 Sample1_Rep2 Sample2_Rep2
Sample1_Rep3 Sample2_Rep3
## 2 WASH7P function=calcium_channel 0.436872902664438 19.3481066900161
9.42244941532453 12.0087099778137 11.9884205794826 20.8571149438669
11.3846931989651
## 10 AL627309.1 function=leucine_rich_repeat 0.841183956745979 3.22468444833601
2.69212840437844 1.20087099778137 0 0.869046455994453 1.13846931989651
```

```
## 13 AL627309.6 function=zinc_finger 0.897413389062224 8.06171112084003
12.114577819703 9.60696798225094 10.898564163166 14.7737897519057 5.69234659948257
## 14 AL627309.7 function=dna_binding 0.933805662095901 11.286395569176
18.8448988306491 10.2074034811416 6.53913849789957 14.7737897519057
13.6616318387582
## 20 FO538757.1 function=transcription_factor 0.992060197430069 31.4406733712761
24.2291556394059 6.60479048779752 14.1681334121157 11.2976039279279
11.3846931989651
## 22 AP006222.1 function=protein_kinase 0.429110680613957 3.22468444833601
13.4606420218922 4.80348399112547 7.62899491421617 2.60713936798336
3.41540795968954
## 25 AL732372.2 function=calcium_channel 0.496779196348994 10.480224457092
18.8448988306491 12.6091454767044 11.9884205794826 11.2976039279279
19.3539784382407
## 31 AL669831.3 function=leucine_rich_repeat 0.975043997893164 2.41851333625201
2.69212840437844 1.80130649667205 1.0898564163166 0.869046455994453
1.13846931989651
## 51 LINC01128 function=zinc_finger 0.751825089638109 130.599720157609
131.914291814543 105.67664780476 101.356646717443 128.618875487179
99.0468308309967
## 61 NOC2L function=dna_binding 0.582090218088542 2125.06705145343
2449.83684798438 1996.44803381152 2170.99398130266 1972.73545510741
1958.167230222
## 62 KLHL17 function=transcription_factor 0.857502880727654 170.908275761809
164.219832667085 125.491019268153 151.490041868007 143.392665239085
95.6314228713071
## 63 PLEKHN1 function=protein_kinase 0.406428109009174 81.4232823204843
71.3414027160286 65.4474693790845 59.9421028974128 75.6070416715175
46.677242115757
## 66 HES4 function=calcium_channel 0.821788994524869 49.9826089492082
30.959476650352 46.8339689134733 73.0203798932119 42.5832763437282
19.3539784382407
## 67 ISG15 function=leucine_rich_repeat 9.88450644059165e-11 292.640113686493
75.3795953225963 210.152424611739 73.0203798932119 179.892616390852
81.969791032549
## 70 AGRN function=zinc_finger 0.593225572130678 3073.9304503763 3468.80744904162
1880.56398252562 2711.56276379569 1958.8307118115 1869.36662327007
## 72 AL390719.1 function=dna_binding 0.479053739994787 32.2468444833601
18.8448988306491 31.8230814412062 23.9768411589651 22.5952078558558
21.6309170780338
## 74 RNF223 function=transcription_factor 0.464818611236974 21.7666200262681
8.07638521313531 5.40391949001615 3.26956924894979 4.34523227997227
4.55387727958605
## 75 C1orf159 function=protein_kinase 0.914057189984315 164.458906865137
110.377264579516 149.50843922378 188.545160022771 163.380733726957
```

```
160.524174105408
## 81 AL390719.2 function=calcium_channel 0.872925529430625 6.44936889667203
6.7303210109461 6.00435498890683 10.898564163166 6.95237164795563 4.55387727958605
## 84 TNFRSF18 function=leucine_rich_repeat 0.85812998347552 1.61234222416801
2.69212840437844 4.20304849223478 4.35942566526638 3.47618582397781
0
## 86 SDF4 function=zinc_finger 0.966065399280551 983.528756742484
1058.00646292073 840.609698446957 966.70264127282 899.463081954259
729.758834053665
## 87 B3GALT6 function=dna_binding 0.0118637218734694 119.313324588432
218.062400754653 133.897116252622 187.455303606454 142.52361878309
165.078051384994
## 90 UBE2J2 function=transcription_factor 0.942646896336212 628.813467425523
663.609651679285 680.893855742035 731.293655348436 776.927531659041
656.896797580288
## 91 LINC01786 function=protein_kinase 0.481166894671379 9.67405334500804
2.69212840437844 4.80348399112547 4.35942566526638 5.21427873596672
2.27693863979303
```

It is a little hard to read because the tabs are not lined up. Here is a trick to make it more readable.

```
<expression.txt column -t|head -25


## Row    Gene                  Function                       padj                  Sampl
## 2      WASH7P                function=calcium_channel       0.436872902664438     19.34
## 10     AL627309.1            function=leucine_rich_repeat   0.841183956745979     3.224
## 13     AL627309.6            function=zinc_finger           0.897413389062224     8.061
## 14     AL627309.7            function=dna_binding           0.933805662095901     11.28
## 20     FO538757.1            function=transcription_factor  0.992060197430069     31.44
## 22     AP006222.1            function=protein_kinase        0.429110680613957     3.224
## 25     AL732372.2            function=calcium_channel       0.496779196348994     10.48
## 31     AL669831.3            function=leucine_rich_repeat   0.975043997893164     2.418
## 51     LINC01128             function=zinc_finger           0.751825089638109     130.5
## 61     NOC2L                 function=dna_binding           0.582090218088542     2125.
## 62     KLHL17                function=transcription_factor  0.857502880727654     170.9
## 63     PLEKHN1               function=protein_kinase        0.406428109009174     81.42
## 66     HES4                  function=calcium_channel       0.821788994524869     49.98
## 67     ISG15                 function=leucine_rich_repeat   9.88450644059165e-11  292.6
## 70     AGRN                  function=zinc_finger           0.593225572130678     3073.
## 72     AL390719.1            function=dna_binding           0.479053739994787     32.24
## 74     RNF223                function=transcription_factor  0.464818611236974     21.76
## 75     C1orf159              function=protein_kinase        0.914057189984315     164.4
## 81     AL390719.2            function=calcium_channel       0.872925529430625     6.449
```

```
## 84     TNFRSF18               function=leucine_rich_repeat   0.85812998347552     1.612
## 86     SDF4                   function=zinc_finger           0.966065399280551    983.5
## 87     B3GALT6                function=dna_binding           0.0118637218734694   119.3
## 90     UBE2J2                 function=transcription_factor  0.942646896336212    628.8
## 91     LINC01786              function=protein_kinase        0.481166894671379    9.674
```

How many genes are in your file?

```
# You count the number of lines in the file
# but don't forget to remove 1 because of the header line

wc -l expression.txt

# Or you could use grep remove the header line
#then pipe the result into the wc command

grep -v "col_" expression.txt | wc -l

# Or you can use tail to print all but the first line
#(the plus tells it to start on the 2nd line)

tail -n +2 expression.txt | wc -l


## 14927 expression.txt
## 14927
## 14926
```

Use awk to manipulate the columns. We want to remove the first column since it is just the row number. We'll practice reordering columns as well.

```
# Reorder the columns so that replicates from a single sample are together
# Note: The backslashes at the end of the lines allow the command
# to be continued on the next lines. I wrote it this way so it
# wouldn't run off the page. If you prefer to write it on one line,
# don't echo the backslashes or the comments (the # and everything after).
# Note: "\t" = tab

awk '{print $2 "\t" $3 "\t" $4 \
$5 "\t" $7 "\t" $9 \
$6 "\t" $8 "\t" $10}' \
expression.txt > reorder.expression.txt
```

- Starting with the original file, how would you:
  - Make a new file that has the gene information and sample 1 only.
  - Make a new file that has all the samples with the gene information (cols 2&3) at the end
  - Make a new file that has the gene information (cols 2&3) both at the beginning and at the end

  Hint: don't forget to use the up arrow to get the previous command so you can edit it rather than retyping it.

Use the up arrow to find the command you used to reorder the file. Change the command so there is an underscore between $2 and $3 rather than a tab. Change the output file to combine.expression.txt.

```
awk '{print $2 "_" $3 "\t" \
$5 "\t" $7 "\t" $9 \
$6 "\t" $8 "\t" $10}' \
expression.txt > combine.expression.txt
```

The col_ on each of the headers is kind of annoying. Let's remove it. We'll use vim and manually edit the document. Here is a vim cheatsheet (https://devhints.io/vim) but hints are also below.

1. Open the file in vim (vim combine.expression.txt).
2. Go into insert mode (i).
3. Manually remove the col_ from each column.
4. Exit insert mode (esc)
5. Save/write and quit (:wq)

Let's also clean up the gene names to make them shorter. We'll remove "function=". We'll use sed.

```
# Here we find "function=" and replace it with nothing.
# If we wanted to replace "function=" with something else,
# we would put that something else between the last two forward slashes.

sed 's/function=//' combine.expression.txt > final.expression.txt
```

- Head the final file to make sure it looks like you expect.
- Remove all of the files except the one you just made ("final.expression.txt").

```
## Gene_Function Sample1_Rep1 Sample1_Rep2 Sample1_Rep3Sample2_Rep1
Sample2_Rep2 Sample2_Rep3
## WASH7P_calcium_channel 19.3481066900161 12.0087099778137 20.85711494386699.42244941532453
11.9884205794826 11.3846931989651
## AL627309.1_leucine_rich_repeat 3.22468444833601 1.20087099778137
0.8690464559944532.69212840437844 0 1.13846931989651
## AL627309.6_zinc_finger 8.06171112084003 9.60696798225094 14.773789751905712.114577819703
10.898564163166 5.69234659948257
## AL627309.7_dna_binding 11.286395569176 10.2074034811416 14.773789751905718.8448988306491
6.53913849789957 13.6616318387582
## FO538757.1_transcription_factor 31.4406733712761 6.60479048779752
11.297603927927924.2291556394059 14.1681334121157 11.3846931989651
## AP006222.1_protein_kinase 3.22468444833601 4.80348399112547 2.6071393679833613.4606420218
7.62899491421617 3.41540795968954
## AL732372.2_calcium_channel 10.480224457092 12.6091454767044 11.297603927927918.8448988306
11.9884205794826 19.3539784382407
## AL669831.3_leucine_rich_repeat 2.41851333625201 1.80130649667205
0.8690464559944532.69212840437844 1.0898564163166 1.13846931989651
## LINC01128_zinc_finger 130.599720157609 105.67664780476 128.618875487179131.914291814543
101.356646717443 99.0468308309967
```

One last thing. Let's look for genes that are significantly different between the two samples. We'll use p-adjusted (column 4) <=0.05 as our cutoff. This means that the gene expression between the two samples is different enough that there is a 5% chance or less that difference is due to some weird artifact with the sampling and a 95% chance that the difference is biologically meaningful.

```
# First we'll look at it.
awk '$4<=0.05{print}' expression.txt | head

# Then let's count them.
awk '$4<=0.05{print}' expression.txt | wc -l
```

```
## 67 ISG15 function=leucine_rich_repeat 9.88450644059165e-11 292.640113686493
75.3795953225963 210.152424611739 73.0203798932119 179.892616390852
81.969791032549
## 87 B3GALT6 function=dna_binding 0.0118637218734694 119.313324588432
218.062400754653 133.897116252622 187.455303606454 142.52361878309
165.078051384994
## 92 SCNN1D function=calcium_channel 0.0455728745148753 23.3789622504361
13.4606420218922 19.2139359645019 3.26956924894979 27.8094865918225
10.2462238790686
## 121 SSU72 function=leucine_rich_repeat 6.03680933947113e-05 1271.33184375647
903.209079668966 1178.05444882352 875.154702302226 1188.85555180041
939.237188914624
```

```
## 126 MIB2 function=protein_kinase 0.012594759266975 209.604489141841
380.936169219549 147.106697228217 262.6553963323 130.356968399168 188.985907102821
## 146 PRKCZ function=leucine_rich_repeat 0.00159620924533322 211.216831366009
327.09360113198 224.562876585116 350.933766053944 232.904450206514
281.201922014439
## 194 TP73-AS1 function=leucine_rich_repeat 0.00270321866739159 231.371109168109
149.413126443003 217.958086097318 152.579898284323 202.487824246708
149.139480906443
## 198 LRRC47 function=zinc_finger 0.0310575351232366 522.398880630434
732.258925990935 480.948834611437 682.250116614189 521.427873596672
565.819251988567
## 203 C1orf174 function=protein_kinase 0.0449186081911594 394.217673809078
487.275241192497 339.246056873236 462.099120518236 353.701907589743
409.848955162745
## 239 HES2 function=dna_binding 0.0298327147440432 191.868724675993
26.9212840437844 65.4474693790845 40.324687403714 59.9642054636173
10.2462238790686
## 2527
```

- How many genes have a p-adjusted value <= 0.01?

```
## 1627
```

You made it to the end! Congratulations!

Questions?

If you have extra time, try it again without looking at the commands I gave you or try altering the commands to ask different questions.